THE USE OF INFORMATION THEORY FUNCTIONALS FOR DATA SMOOTHING

N.S. Tzannes

Department of Electrical Engineering, University of Patras, Patras, Greece

S. Politaki

3 Alatsaton St., Athens, Greece

Abstract. Recent closed form solutions to the Mutual Information Principle (MIP), are used in reconstituting signals on the basis of limited a priori information about them. Most emphasis is given to everywhere positive signals which must be optimally smoothed using a few measured values obtained with an instrument of known average error. The method is compared to results obtained from other classical methods, such as Least Squares, Lagrange and Newton.

Keywords. Approximation theory; data processing; filtering; function approximation; signal processing; smoothing.

1. PRELIMINARY BACKGROUND

The Mutual Information Principle (MIP) was introduced by Tzannes and Noonan (1973) as a method for picking a prior pdf of a continuous random variable (r.v.), utilizing data gleaned from experiment. They argued that as measurements are inevitably discrete, the observations - outputs of some instrument - define a discrete r.v. whose pmf can be found using established methods such as the Maximum Entropy Principle (MEP). The pdf of the continuous r.v. - input to the instrument - is then obtained by minimizing the mutual information between the continuous-input and the discrete-output of the instrument (channel) subject to various constraints that safeguard that the conditionals are legitimate pdf's, as well as a constraint that reflects the error (or noise) of the instrument.

Mathematically, the MIP proceeds as follows. Let us denote the continuous r.v. whose pdf is sought by X, and the observed discrete one by Y. The two are related via an instrument, and this relation is expressed in the form of an average error

$$E\{d(X, Y)\} \leq D \qquad (1.1)$$

where both $d(X, Y)$ and D are known a priori.

Let us next assume that the known pmf of Y is denoted by $q(y_k)$. Then a pdf for X that minimizes the mutual information between X and Y and satisfies all the constraints is given by (Cyranski and Tzannes, 1983)

$$f(x) = A \sum_k q(y_k)\, e^{-sd(x-y_k)} \qquad (1.2)$$

If $d(x,y)$ is specified to be the usual case of,

$$d(X, Y) = (X - Y)^2 \qquad (1.3)$$

then the above pdf becomes

$$f(x) = \sum_k q(y_k) \frac{e^{-(x-y_k)^2/2D}}{\sqrt{2\pi D}} \qquad (1.4)$$

and similar expressions can be found for other error functions. The above solution is not unique. It is, however, a solution, and a pleasing one as it indicates a mechanism controlled by $d(X, Y)$ smoothing out the discrete pmf to a continuous pdf.

The MIP is, of course, controversial, but it addresses a problem for which no solutions are yet fully accepted. We shall attempt to apply it here to the age-old problem of smoothing out observed discrete data, to an uninterrupted continuous function.

2. DATA SMOOTHING

Let us assume that measurements with a noisy instrument have provided us with values $y(t_k)$ $(k = 1, \ldots, n)$. Our aim is to use the MIP to arrive at a smooth curve one that presumably represents in an "optimum" way the underlying function $x(t)$, which is at the input to the noisy instrument. The parameter t may be various things - not just time.

To begin with, if the MIP is to be used, $y(t_k)$ must behave like a pmf. This means that,

$$0 \leq y(t_k) \leq 1 \quad \forall k \qquad (2.1)$$

and

$$\sum_k y(t_k) = 1 \qquad (2.2)$$

conditions which are not normally met in practice. To correct the situation we can proceed as follows.

(a) To meet (2.1) we add a constant to all $y(t_k)$ so that the biggest negative value becomes zero. This shifts everything upward. We must remember, at the end, to shift the solution downward by the same amount.

(b) Once (a) is accomplished, (2.2) can be met by dividing each $y(t_k)$ by their sum - a normalizing procedure. The obtained smooth curve, will, of course, be normalized in the same manner.

Let us assume that (a) and (b) have been done, and the values $y(t_k)$ meet both conditions (2.1) and (2.2). Next we assume that the instrument has been studied and its average "noise" effects are known. For simplicity in the presentation here, we assume that the input-output effects of the instruments are given by a mean-squared error criterion of the form discussed earlier. All this leads to a solution for x(t) of the form (1.4) i.e.

$$x(t) = \sum_k y(t_k)\frac{e^{-(t-t_k)^2/2D}}{\sqrt{2\pi D}} \qquad (2.3)$$

where D is the mean-squared error of the instrument. There are some unresolved theoretical questions about all this, of course, which are described in some detail elsewhere (Tzannes and Avgeris, 1980). Here we will try to use it and compare the result with some other known classical smoothing methods.

3. THE APPLICATION

To apply the above theory to a specific example and see how it comes out, we chose the function,

$$f(t) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\} \qquad (3.1)$$

the well known normalized Gaussian curve, whose values are tabulated and given in most texts on mathematical statistics.

Let us assume that an instrument whose average error is $D=0.1$, has given us the following values $f(t_k)$ for this function,

k	t_k	$f(t_k)$
0	-1	0.2420
1	-0.5	0.3521
2	0	0.3989
3	0.5	0.3521
4	1	0.2420
5	1.5	0.1295
6	2	0.0540

(3.2)

It should be noted that the only information that we have about the "noise" is the average error D. No further assumptions such as the nature of its pdfs, its power spectral density, etc. are needed for the application of the MIP approach. Of course, in an actual application the original f(t) is not known. Here we purposely assume it known, in order to check our final results and see how the various methods approximate it.

The experimentally obtained $f(t_k)$'s of (3.2) were smoothed to a continuous function by using three classical methods, and the MIP. We omit the calculations and go directly to the results.

Applying Newton's method we obtained the continuous function,

$$\begin{aligned} f_{NEW}(t) = {} & 0.2420 + 0.2202(t+1) - 0.1266(t+1)\cdot \\ & \cdot(t+0.5) - 0.0404(t+1)(t+0.5)t + \\ & + 0.0404(t+1)(t+0.5)t(t-0.5) - \\ & - 0.0080(t+1)(t+0.5)t(t-0.5)\cdot \\ & \cdot(t-1) - 0.0019(t+1)(t+0.5)t\cdot \\ & \cdot(t-0.5)(t-1)(t-1.5) \end{aligned}$$

Next we tried Lagrange's method. The final result is,

$$\begin{aligned} f_{LAG}(t) = {} & 0.0215(t+0.5)t(t-0.5)(t-1)\cdot \\ & \cdot(t-1.5)(t-2) - 0.1877(t+1)t\cdot \\ & \cdot(t-0.5)(t-1)(t-1.5)(t-2) + \\ & + 0.5318(t+1)(t+0.5)(t-0.5)\cdot \\ & \cdot(t-1)(t-1.5)(t-2) - \\ & - 0.6259(t+1)(t+0.5)t(t-1)\cdot \\ & \cdot(t-1.5)(t-2) + 0.3226(t+1)\cdot \\ & \cdot(t+0.5)t(t-0.5)(t-1.5)(t-2) - \\ & - 0.0690(t+1)(t+0.5)t(t-0.5)\cdot \\ & \cdot(t-1)(t-2) + 0.0034(t+1)\cdot \\ & \cdot(t+0.5)t(t-0.5)(t-1)(t-1.5) \end{aligned}$$

The MIP approach led to,

$$\begin{aligned} f_{MIP}(t) = {} & 0.1367\, e^{-(t+1)^2/0.2} + \\ & + 0.1988\, e^{-(t+0.5)^2/0.2} + \\ & + 0.2253\, e^{-t^2/0.2} + \\ & + 0.1988\, e^{-(t-0.5)^2/0.2} + \\ & + 0.1367\, e^{-(t-1)^2/0.2} + \\ & + 0.0732\, e^{-(t-1.5)^2/0.2} + \\ & + 0.0305\, e^{-(t-2)^2/0.2} \end{aligned}$$

Our final effort involves the method of Least Squares. Here we "stacked the deck", and picked the elements of the expansion to be Gaussian functions of variance equal to D, i.e. elements which come from the solution of the MIP. The final result is

$$\begin{aligned} f_{LS}(t) = {} & 0.1760\ e^{-(t+1)^2/0.2} + \\ & + 0.2235\ e^{-(t+0.5)^2/0.2} \\ & + 0.2667\ e^{-t^2/0.2} + \\ & + 0.2298\ e^{-(t-0.5)^2/0.2} + \\ & + 0.1525\ e^{-(t-1)^2/0.2} + \\ & + 0.0751\ e^{-(t-1.5)^2/0.2} + \\ & + 0.0313\ e^{-t(t-2)^2/0.2} \end{aligned}$$

Which of the above four methods "best" approximates the original underlying f(t) which led to the experimentally obtained values of (3.2) ? Generally speaking, the answer to this question is not available as the underlying degree of optimality differs from method to method. Nevertheless, in view of our assumed known, original f(t) we can get an intuitive notion of the answer by finding the "distance" (or error),

$$\int_{-\infty}^{+\infty} [f(t) - \hat{f}(t)]^2\, dt \qquad (3.3)$$

where f(t) the function of (3.1) and $\hat{f}(t)$ the result of each method attempted. This distance again favors the method of Least Squares, so the results of the comparison is somewhat suspected a priori.

Computer results lead to the following errors for each method.

(a) Newton: Error = 11.0955

(b) Lagrange: Error = 2.8867

(c) Least Squares: Error = 0.0443

(d) MIP: Error = 0.1100

It is rather interesting that the MIP did as well as that, even though the overall approach was designed to favor the method of Least Squares.

4. CONCLUSION

A first attempt at using the MIP method as a smoothing filter appears to be promising. More work needs to be done to theoretically compare the method to other existing classical methods and ascertain its effectiveness.

REFERENCES

Cyranski, J.F., and N.S. Tzannes (1983). Some closed form solutions to the Mutual information principle. Kybernetes to be published.

Tzannes, N.S., and T.G. Avgeris (1980). An information theory approach to the approximation of functions. In D.G. Lainiotis and N.S. Tzannes (Editors), Advances in Communications. D. Reidel Publ. Corp., Dordrecht, Holland.

Tzannes, N.S., and J.P. Noonan (1973). The mutual information principle and applications. Information and Control, 22, 1-12.